ISSN: 2321-7782 (Online)
Volume 3, Issue 11, November 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

B-PSO Swarm Search for ECG Data Mining and Its Classification for Atrial Fibrillation

Ghuge Sharad Tryambak[1]
Dept.of computer Engg.
SND College of Engg. & RC
Yeola Dist Nashik (MH India)

Prof. Rokade P.P[2]
Dept. of Computer Engg.
SND College of Engg. & RC
Yeola Dist Nashik (MH India)

Abstract: Particle Swarm Optimization (PSO) is a biologically inspired computational search and optimization method developed in 1995 by Eberhart and Kennedy based on the social behavior of birds flocking or fish schooling. A number of basic variations have been developed due to improve speed of convergence and quality of solution found by the PSO. On the other hand, basic PSO is more appropriate to process static, simple optimization problem. Modification PSO is developed for solving the basic PSO problem. The observation and review related studies in the period between 2002 and 2010 focusing on function of PSO, advantages and disadvantages of PSO, the basic variant of PSO, Modification of PSO and applications that have implemented using PSO. The application can show which one the modified or variant PSO that haven't been made and which one the modified or variant PSO that will be developed. Big Data though it is a hype up-springing many technical challenges that confront both academic research communities and commercial IT deployment, the root sources of Big Data are founded on data streams and the curse of dimensionality. It is generally known that data which are sourced from data streams accumulate continuously making traditional batch-based model induction algorithms infeasible for real-time data mining. Feature selection has been popularly used to lighten the processing load in inducing a data mining model. However, when it comes to mining over high dimensional data the search space from which an optimal feature subset is derived grows exponentially in size, leading to an intractable demand in computation. In order to tackle this problem which is mainly based on the high-dimensionality and streaming format of data feeds in Big Data, a novel lightweight feature selection is proposed. The feature selection is designed particularly for mining streaming data on the fly, by using accelerated particle swarm optimization (APSO) type of swarm search that achieves enhanced analytical accuracy within reasonable processing time. In this paper, a collection of Big Data with exceptionally large degree of dimensionality are put under test of our new feature selection algorithm for performance evaluation.

Keywords: Feature Selection, Swarm Intelligence, Classification, Big Data, Particle Swarm Optimization

I. INTRODUCTION

Theory of Particle Swarm Optimization (PSO) has been growing rapidly. PSO has been used by many applications of several problems. The algorithm of PSO emulates from behaviour of animals societies that don't have any leader in their group or swarm, such as bird flocking and fish schooling. Typically, a flock of animals that have no leaders will find food by random, follow one of the members of the group that has the closest position with a food source (potential solution). The flocks achieve their best condition simultaneously through communication among members who already have a better situation. Animal which has a better condition will inform it to its flocks and the others will move simultaneously to that place. This would happen repeatedly until the best conditions or a food source discovered. The process of PSO algorithm in finding optimal values follows the work of this animal society. Particle swarm optimization consists of a swarm of particles, where particle represent a potential solution.

1.1 Particle swarm optimization (PSO)

Particle swarm optimization is an optimization method which works based on the movement of bird flocking or fish schooling [1]. In PSO, all particles in a search space find their best solution by adjusting their own flying experience or the other social flying experience. Instead of current position and current velocities, each particle modifies and keeps track of their position based on personal best (p-best) and global best (g-best). Later on, the inertia weight is used in order to balance between the local and global search. A large inertia weight facilitates global search while the small inertia weight facilitates the local search. The value of the inertia weight is linearly decreased from large to small throughout the iteration in order to have more global search ability at the beginning and more local search ability near the end of iteration [2]. However, in this research, PSO with BPSO is integrated for more optimization.

II. Existing System

There are traditional methods available for the feature extraction. However it is recently reported that many proposed methods are limited to one or more of the following constraints in their designs

1. The size of the resultant feature set is assumed fixed. Users are required to explicitly specify the maximum dimension for feature subset. Although the number of combinations reduces from 2k to k!/(k − s)! s! where there is a maximum of k features in the original dataset and s is the upper limit of the subset (for s≤k), the major drawback is that users may not know in advance what would be the ideal size of s [3].
2. The feature becomes minimal. By the principle of removing redundancy, the feature set may shrink to its most minimal size.
3. The feature selection methods are custom designed for some particular classifier and optimizer. Although an exhaustive computing method may be used for finding the most appropriate feature subset, this is quite impractical for data streams which are usually in very high dimensions and their amount may accumulate to infinity.

There is large amount of heart related data present, which is in unstructured format. Hence by analysing the data and formatting it into structured manner helps for making the decision. For diagnosing the disease there are many ways in which heart related diseases can be diagnosed and treatment can be provided. Different approaches have different aspects in diagnosing the diseases. By using the Neural network approach the accuracy secured was around 80- 90% but the hidden layers description cannot be evaluated [5]. In fuzzy logic approach the weighted rules are generated initially and then the fuzzy rule decision is provided and the accuracy obtained id around 79.05%. In naive bayes classification approach helps in predicting whether the patient is prone to heart disease or not and depicting the risk factor for heart attack. The accuracy observed for naive bayes approach was around 90%. Similarly by using Support vector machines concept the accuracy was achieved around 80% approx. While as by using particle swarm approach the accuracy is increased.

III. Proposed system

We are going to analyse the following algorithms and check out their efficiency and then comment on that which method is better whether traditional approach or the incremental approach. The algorithms are

1. PSO
2. BPSO
3. Random Forests
4. K-Means

The dataset which we are getting the firstly get pre-processed from where Data Cleaning and Conversion processes are done. After that the next thing is that the feature extraction is done through the PSO [4] and BPSO method. This algorithm leads to process the big data. Then data mining will be done which will be produced the classified output of the data set. After this the Random forests algorithm and k-means algorithms are used on training database which produce the final output.

Here the proposed system is taking input as a text file of voltage verses time graph points which varies from -1 to +1 and these are pre-processed by applying low pass filter and high pass filter. This filtered data is input to the feature extraction algorithms K-means And Random Forest.

IV. System Architecture

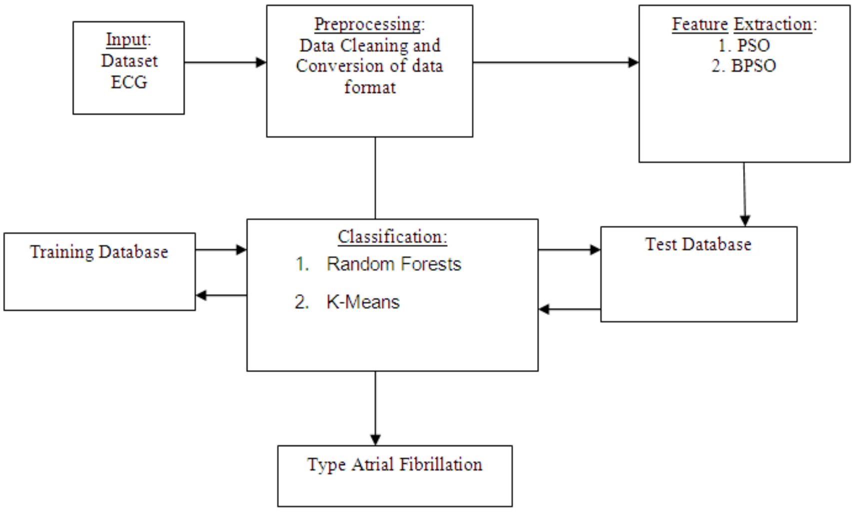


Fig.1 Architecture of the Proposed System

V. Working Of The Proposed System

1. Data Conversion

The Input Dataset is converted to the format required by the System. The System can input data in the format of ARFF or CVV. The data is cleaned and properly formatted to be used by the system.

2. Feature Extraction Normalization Process

The Features of the dataset is extracted and the data is optimized by this module. Here we are using PSO (Particle Swarm Optimization and B-Positive PSO)

Dataset Data though it is a manageable many technical challenges that confront both academic research communities and commercial IT deployment, the root sources of Data are founded on data streams and the curse of dimensionality. It is generally known that data which are sourced from data streams accumulate continuously making traditional batch-based model induction algorithms infeasible for real-time data mining. Feature selection has been popularly used to lighten the processing load in inducing a data mining model. However, when it comes to mining over high dimensional data the search space from which an optimal feature subset is derived grows exponentially in size, leading to an intractable demand in computation. In order to tackle this problem which is mainly based on the high-dimensionality and streaming format of data feeds in Data mining, a need of lightweight feature selection is proposed. The Input data features will be selected by the PSO/B-PSO technique and will then input to the Data mining for training and classification.

3. *PSO Optimization*

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by social behaviour of bird flocking or fish schooling. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles [5].

Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called pbest. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbours of the particle. This location is called lbest. when a particle takes all the population as its topological neighbours, the best value is a global best and is called gbest [6].

The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its best and best locations (local version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward p-best and best locations.

The classical PSO algorithm is as follows;

1. Initialize the particles with random velocities and positions in a given Dimension.
2. Compute the fitness of all particles using the desired benchmark function, choose the lowest one as global best and assign the current positions of all particles as their best.
3. Calculate the next velocities and positions using the equations 1 and 2.
4. Calculate the fitness of all particles using the updated velocities and positions. If the new fitness is less than the particles best fitness, the new fitness is considered the best one. In the same way the new position is considered the particle best position.
5. In the same way if the new particle's best fitness is less than the overall global best fitness, it is considered the new best global fitness and its corresponding position is considered new global best position.
6. Repeat third step for desired number of function evaluations.

The mathematical equations [2], [3] used to calculate the new velocity and position are as follows.

$$\vec{v}_i(j+1)=\vec{v}_i(j)+\vec{r}_i(j)(\vec{p}_i(j)-\vec{x}_i(j))+\vec{r}_i(j)(\vec{g}(j)-\vec{x}_i(j)) \quad (1)$$

$$\vec{x}_i(j+1)=\vec{x}_i(j)+\vec{v}_i(j+1) \quad (2)$$

Where, $\vec{v}_i(j)$ is the velocity vector of particle i at iteration j

$\vec{x}_i(j)$ is the position vector of particle i at iteration j ,

$\vec{p}_i(j)$ is the D-dimensional personal best of particle i ,

$\vec{g}_i(j)$ is the D-dimensional global best of the whole swarm.

4. B-PSO Optimization

The modifications made in the PSO in order to improve the output values While initializing the B-PSO the velocity (which is actually the displacement of the particles) is clamped to be from zero to a specific percentage (30%) of the total search space. The clamping effect is already a part of Standard PSO but in this case the whole negative part of the velocity is removed which makes negative displacement impossible. Overall displacement of the particles in a two dimensional search space is shown below. It can be noticed that the particles only move in positive direction as the lower bound of velocity (displacement) vector is kept 0.

In Standard PSO, every time a new velocity is calculated using

$$\vec{v}_i(j+1) = \underbrace{w\,\vec{v}_i(j)}_{\text{inertia part}} + \underbrace{c_1\vec{r}_i(j)(\vec{p}_i(j) - \vec{x}_i(j))}_{\text{cognition part}} + \underbrace{c_2\vec{r}_i(j)(\vec{g}(j) - \vec{x}_i(j))}_{\text{social part}}$$

It is simply added to the position of a particle using

$$\vec{x}_i(j+1) = \vec{x}_i(j) + \vec{v}_i(j+1)$$

Where, $\vec{v}_i(j)$ is the velocity vector of particle i at iteration j

$\vec{x}_i(j)$ is the position vector of particle i at iteration j ,

$\vec{p}_i(j)$ is the D-dimensional personal best of particle i ,

$\vec{g}_i(j)$ is the D-dimensional global best of the whole swarm.

To get the new position, In the B.P.S.O the velocity is multiplied with R, where R is a uniform random number generator which generates random numbers between (0, 1).

$$\vec{x}_i(j+1) = \vec{x}_i(j) + R(0,1)\vec{v}_i(j+1)$$

In Standard PSO, if the new position of a particle is less than the particles optimum position, the new position becomes the optimum one otherwise nothing is done. In the B.P.S.O, instead of doing nothing in the second case, the particle is thrown randomly around its best available position for future evaluations.

$$\textit{if} \quad f(\vec{x}_i(j)) > f(\vec{p}(j)) \quad \textit{then} \quad \vec{x}_i(j) = R(0,1)\vec{p}_i(j)$$

5. Training Module

The Normalized data (after applying Feature extraction PSO/BPSO) is given as a Input to the KMEANS and Random Forest Data mining Algorithm for Training.

6. Classification

The Classification requires Cleaning and Optimization of the Test data input. The Classification refers to the trained dataset and accordingly classifies the output and detects the Atrial Fibrillation.

7. *Graphical Analysis*

Analysis of classification along with feature selection will be shown as output.

A. Analysis of Data Optimization using PSO
B. Analysis of Data Optimization using BPSO
C. Classification and Accuracy Result with PSO
D. Classification and Accuracy Result with BPSO

The final result will be shown in the form of graphical layout as discussed above.

VI. Applications

Proposed system is applicable for the heart disease diagnosis by mining and classifying data from existing ECG database and matching present ECG report to the existing data and disease detected.

VII. Conclusion

The process of PSO algorithm in finding optimal values follows the work of an animal society which has no leader. Particle swarm optimization consists of a swarm of particles, where particle represent a potential solution (better condition). Particle will move through a multidimensional search space to find the best position in that space (the best position may possible to the maximum or minimum values). In this paper, I have made review of the different methods of PSO algorithm. Basic particle swarm optimization has advantages and disadvantages, to overcome the lack of PSO. There are several basic variant of PSO. The basic variants as mentioned above have supported controlling the velocity and the stable convergence. At the other hands, modified variant PSO help the PSO to process other conditions that cannot be solved by the basic PSO. The observation and review are to be made to show the absolute function of PSO, advantages and disadvantages of PSO, the basic variant of PSO, Modification of PSO and applications that have implemented using PSO. The application can show which one the modified or variant PSO that haven't been made and which one the modified or variant PSO that will be developed.

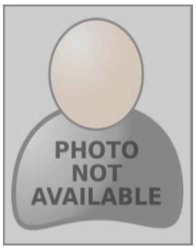
References

1. Simon Fong, Raymond Wong, and Athanasios V. Vasilakos, Senior Member, IEEE "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data"(2015)
2. Quinlan, J. R., C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993
3. Fong, S., Liang, J., Wong, R., Ghanavati, M., "A novel feature selection by clustering coefficients of variations", 2014 Ninth International Con-ference on Digital Information Management (ICDIM), Sept. 29, 2014, pp.205-213
4. Mohamed Medhat Gaber, Arkady Zaslavsky, Shonali Krishnaswamy, "Mining data streams: a review", ACM SIGMOD Record, Volume 34 Issue 2, June 2005, pp.18-26
5. Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. Paper presented at the Micro Machine and Human Science, 1995. MHS'95.,Proceedings of the Sixth International Symposium on.
6. Kennedy, J., & Eberhart, R. 1995. Particle swarm optimization. Paper presented at the Neural Networks, 1995. Proceedings.,IEEE International Conference on.
7. Doctor, S., Venayagamoorthy, G. K., & Gudise, V. G. (2004). Optimal PSO for collective robotic search applications. Paper presented at the Evolutionary Computation, 2004. CEC2004. Congress on.
8. Hereford, J. M. (2006). A distributed particle swarm optimization algorithm for swarm robotic applications. Paper presented at the Evolutionary Computation, 2006. CEC 2006. IEEE Congress on.
9. Hereford, J., & Siebold, M., 2008. Multi-Robot Search Using A Physically-Embedded Particle Swarm Optimization. International Journal of Computational Intelligence Research, 4(2), 197-209.
10. Xue, S., Zhang, J., & Zeng, J., 2009. Parallel Asynchronous Control Strategy For Target Search With Swarm Robots. International Journal of Bio-Inspired Computation, 1(3), 151-163.
11. Jatmiko, W., Sekiyama, K., & Fukuda, T. 2006. A pso-based mobile sensor network for odor source localization in dynamic environment: Theory, simulation and measurement. Paper presented at the Evolutionary Computation, 2006.
12. Indre Zliobaite, Albert Bifet, Bernhard Pfahringer, Geoff Holmes, "Ac-tive Learning with Evolving Streaming Data", ECML/PKDD (3) 2011, pp.597-612
13. Pugh, J., Segapelli, L., & Martinoli, A. (2006). Applying aspects of multi-robot search to particle swarm optimization Ant Colony Optimization and Swarm Intelligence (pp. 506-507): Springer.

14. Pugh, J., & Martinoli, A., 2007. Inspiring And Modeling Multi-Robot Search With Particle Swarm Optimization. Paper presented at the Swarm Intelligence Symposium, 2007. SIS 2007. IEEE.
15. Arinto Murdopo, "Distributed Decision Tree Learning for Mining Big Data Streams", Master of Science Thesis, European Master in Distribut-ed Computing, July 2013
16. Robinson, J., Sinton, S., & Rahmat-Samii, Y., 2002. Particle Swarm, Genetic Algorithm, And Their Hybrids: Optimization Of A Profiled Corrugated Horn Antenna. Paper presented at the Antennas and Propagation Society International Symposium, 2002. IEEE.
17. Vesterstrom, J. S., Riget, J., & Krink, T., 2002. Division Of Labor In Particle Swarm Optimisation. Paper presented at the Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on.
18. Li, J., & Xiao, X. (2008). Multi-swarm and multi-best particle swarm optimization algorithm. Paper presented at the Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on.
19. Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. Paper presented at the Evolutionary Computation Proceedings,1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Conference on.
20. Hart, P. E., Nilsson, N. J., & Raphael, B., 1968. A Formal Basis For The Heuristic Determination Of Minimum Cost Paths. Systems Science and Cybernetics, IEEE Transactions on, 4(2), 100-107.

AUTHOR(S) PROFILE

Ghuge Sharad Tryambak, received the Bachelor degree in Information Technology and Pursuing Master degree in Computer Engineering from SND College Of Engineering And Research Center, Yeola (Nashik).



Prof. Rokade P P. Head Of The Department Computer Engineering, SND College of Engineering And Research Center, Yeola. Master of Computer Engineering And Doing Research on Data Mining which is area of Interest and simultaneously Guiding me for my project and Research.